CONDITIONAL & MARGINAL GAUSSIAN DIST.

[1] Introduction

Conditional probability distribution: $p(x \mid d)$

where: $p(x \mid d) = \dfrac{p(x,d)}{p(d)} = \dfrac{p(d \mid x)\, p(x)}{p(d)}$

Marginal distribution: $p(d)$

where: $p(d) = \int p(x,d)\, dx = \sum_{x} p(x,d)$

Motivation: we know that solving $p(d)$ and thus $p(x \mid d)$ is analytically intractable & computationally prohibitively expensive.

$\downarrow$

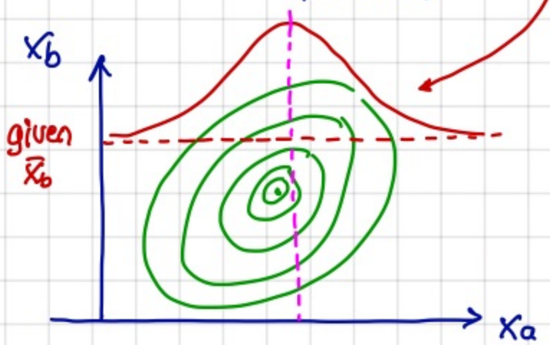
Solution: to use the Gaussian distribution function

$\boxed{\text{Two Cases}}$

case 1:

We want to know the mean & the covariance matrix

$p(\bar{x}_a \mid \bar{x}_b)$



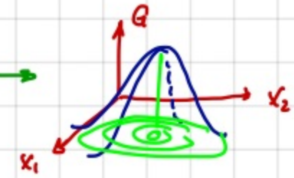
We want to know:

1. $p(\bar{x}_a \mid \bar{x}_b)$
2. $p(\bar{x}_a)$

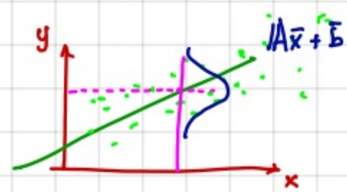
case 2:

$$\underset{D\times 1}{\bar{y}} = \underset{D\times M}{\mathbb{A}}\, \underset{M\times 1}{\bar{x}} + \underset{D\times 1}{\bar{b}}$$

$p(\bar{x}) = G(\bar{x};\, \bar{\mu},\, \Lambda^{-1})$



$p(\bar{y} \mid \bar{x}) = G(\bar{y};\, \mathbb{A}\bar{x} + \bar{b},\, L^{-1})$



We want to know:

1. $p(\bar{x} \mid \bar{y})$
2. $p(\bar{y})$

[2] Conditional & Marginal Gaussian Distributions

CASE 1: $\bar{x} = \begin{pmatrix} \bar{x}_a \\ \bar{x}_b \end{pmatrix}$ and $p(\bar{x}) = G(\bar{x}; \mu, \Sigma)$

$$= G\left(\begin{bmatrix} \bar{x}_a \\ \bar{x}_b \end{bmatrix}; \begin{bmatrix} \bar{\mu}_a \\ \bar{\mu}_b \end{bmatrix}, \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \right)$$

Marginal Gaussian distribution:

$$p(\bar{x}_a) = G(\bar{x}_a; \mu_a, \Sigma_{aa})$$
$$p(\bar{x}_b) = G(\bar{x}_b; \mu_b, \Sigma_{bb})$$

Conditional Gaussian distribution:

$$p(\bar{x}_a | \bar{x}_b) = G(\bar{x}_a; \mu_{a|b}, \Sigma_{a|b}^{-1})$$

where:

$$\mu_{a|b} = \mu_a + \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)$$
$$\Sigma_{a|b} = (\Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba})^{-1}$$

$$p(\bar{x}_b | x_a) = G(\bar{x}_b; \mu_{b|a}, \Sigma_{b|a}^{-1})$$

where:

$$\mu_{b|a} = \mu_b + \Sigma_{ba} \Sigma_{aa}^{-1} (x_a - \mu_a)$$
$$\Sigma_{b|a} = (\Sigma_{bb} - \Sigma_{ba} \Sigma_{aa}^{-1} \Sigma_{ab})^{-1}$$

CASE 2: $\bar{y} = \mathbb{A}\bar{x} + \bar{b}$

where:

$$p(\bar{x}) = G(\bar{x}; \bar{\mu}, \Lambda^{-1}) \rightarrow \text{Prior}$$
$$p(\bar{y}|\bar{x}) = G(\bar{y}; \mathbb{A}\bar{x} + b, L^{-1}) \rightarrow \text{Likelihood}$$

Marginal Gaussian distribution:

$$p(\bar{y}) = G(\bar{y}; \mathbb{A}\bar{\mu} + \bar{b}, L^{-1} + \mathbb{A}\Lambda^{-1}\mathbb{A}^T)$$

Conditional Gaussian distribution:

$$p(\bar{x}|\bar{y}) = G(\bar{x}; \bar{\mu}_{x|y}, \Sigma_{x|y})$$

where:

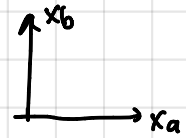
$$\mu_{x|y} = (\Lambda + \mathbb{A}^T L \mathbb{A})^{-1} (\mathbb{A}^T L (\bar{y} - \bar{b}) + \Lambda\bar{\mu})$$
$$\Sigma_{x|y} = (\Lambda + \mathbb{A}^T L \mathbb{A})^{-1}$$

[3] Conditional Gaussian Distributions : Case 1

Let $\bar{x} \in \mathbb{R}^D$: $\bar{x}$ is decomposed to $\bar{x}_a$ and $\bar{x}_b$, meaning : $\bar{x} = \begin{bmatrix} \bar{x}_a \\ \bar{x}_b \end{bmatrix}$

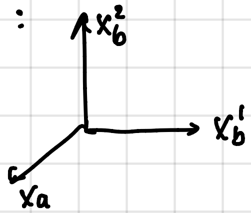
where $\bar{x}_a \in \mathbb{R}^P$ and $\bar{x}_b \in \mathbb{R}^Q$. Thus $D = P + Q$.

Examples : ① $\bar{x} \in \mathbb{R}^2$, then $x_a \in \mathbb{R}^1$ and $x_b \in \mathbb{R}^1$:



② $\bar{x} \in \mathbb{R}^3$, then $x_a \in \mathbb{R}^1$ and $\bar{x}_b \in \mathbb{R}^2$:

$\bar{x}_b = \begin{bmatrix} x_b^1 \\ x_b^2 \end{bmatrix}$



$\Downarrow$

$$p(\bar{x} \mid \bar{\mu}, \Sigma) = p\left(\begin{bmatrix} \bar{x}_a \\ \bar{x}_b \end{bmatrix} \middle| \begin{bmatrix} \bar{\mu}_a \\ \bar{\mu}_b \end{bmatrix}, \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \right)$$

Taking a close look at Σ :

$$\Sigma = \begin{bmatrix} \underset{Q \times P}{\overset{P \times P}{\Sigma_{aa}}} & \overset{P \times Q}{\Sigma_{ab}} \\ \Sigma_{ba} & \underset{Q \times Q}{\Sigma_{bb}} \end{bmatrix} \rightarrow$$

Example :

$$\Sigma = \begin{bmatrix} \boxed{1} & 4 & 5 \\ 4 & 2 & 6 \\ 5 & 6 & 3 \end{bmatrix}$$

$\Sigma_{aa} = 1$

$\Sigma_{bb} = \begin{bmatrix} 2 & 6 \\ 6 & 3 \end{bmatrix}$

$\underset{1 \times 2}{\Sigma_{ab}} = \underset{2 \times 1}{\Sigma_{ba}^T}$

Let's define :

$$\Lambda \equiv \Sigma^{-1}$$

precision matrix

$$= \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$$

$\Downarrow$

Conditional Gaussian distribution :

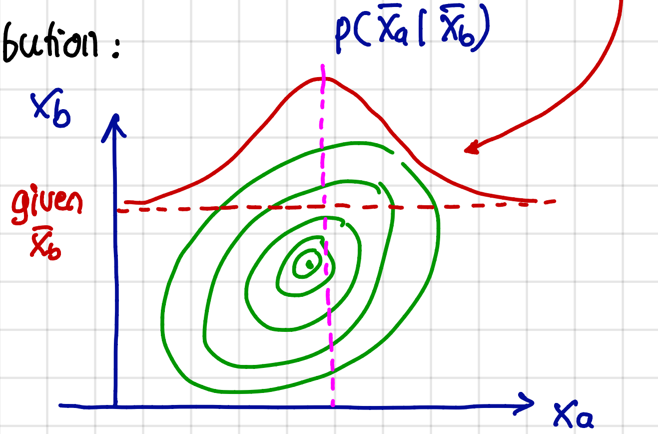
$$p(\bar{x}_a \mid \bar{x}_b) = ?$$

where :

$$p(\bar{x}) = G(\bar{x}; \bar{\mu}, \Sigma)$$

and $\bar{x}_b$ are given.

We want to know the mean & the covariance matrix



[4] Proof : Conditional & Marginal Gaussian Distributions (case 1)

$$p(\bar{x}_a | \bar{x}_b) = ?$$

a Gaussian with certain μ & Σ

We might want to try this : $p(\bar{x}_a | \bar{x}_b) = \dfrac{p(\bar{x}_a, \bar{x}_b)}{p(\bar{x}_b)} = \dfrac{p(\bar{x})}{p(\bar{x}_b)}$

$\to$ it's given

$\to$ we know $\bar{x}_b$, but don't know its probability.

COMPLETE PROOF :

To prove :

① $p(\bar{x}_a) = \int p(\bar{x}_a, \bar{x}_b)\, d\bar{x}_b = G(\bar{x}_a ; \mu_a, \Sigma_{aa})$

$p(\bar{x}_b) = \int p(\bar{x}_a, \bar{x}_b)\, d\bar{x}_a = G(\bar{x}_b ; \mu_b, \Sigma_{bb})$

} Marginal Gaussian distributions

② $p(\bar{x}_a | \bar{x}_b) = G(\bar{x}_a | \bar{\mu}_{a|b}, \Sigma_{a|b})$

$p(\bar{x}_b | \bar{x}_a) = G(\bar{x}_b | \bar{\mu}_{b|a}, \Sigma_{b|a})$

} Conditional Gaussian distributions

PROOF :

Note : $\bar{x} \equiv x$, $\bar{x}_a \equiv x_a$, $\bar{x}_b \equiv x_b$

$$p(x) = p(x_a, x_b) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$
$$= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} Q(x_a, x_b)\right)$$

where :

$$Q(x_a, x_b) = \begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}$$

$\to$ Recall : $\Sigma^{-1} = \Lambda = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$

$$= \begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}^T \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix} \begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}$$

$$= \begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}^T \begin{bmatrix} \Lambda_{aa}(x_a - \mu_a) + \Lambda_{ab}(x_b - \mu_b) \\ \Lambda_{ba}(x_a - \mu_a) + \Lambda_{bb}(x_b - \mu_b) \end{bmatrix}$$

$$= (x_a - \mu_a)^T \Lambda_{aa} (x_a - \mu_a) + (x_a - \mu_a)^T \Lambda_{ab} (x_b - \mu_b)$$
$$+ (x_b - \mu_b)^T \Lambda_{ba} (x_a - \mu_a) + (x_b - \mu_b)^T \Lambda_{bb} (x_b - \mu_b)$$

[•] Inverse of Partitioned Matrix

$$\Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \quad , \quad \Sigma^{-1} = \Lambda = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$$

What is the correlation between Λ_{aa} with Σ's elements ?

$$\mathbb{I} = \Sigma\Sigma^{-1} = \Sigma\Lambda = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}\begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix} = \begin{bmatrix} \Sigma_{aa}\Lambda_{aa} + \Sigma_{ab}\Lambda_{ba} & \Sigma_{aa}\Lambda_{ab} + \Sigma_{ab}\Lambda_{bb} \\ \Sigma_{ba}\Lambda_{aa} + \Sigma_{bb}\Lambda_{ba} & \Sigma_{ba}\Lambda_{ab} + \Sigma_{bb}\Lambda_{bb} \end{bmatrix} = \begin{bmatrix} \mathbb{I} & 0 \\ 0 & \mathbb{I} \end{bmatrix}$$

$$\Downarrow$$

(1) $\Sigma_{aa}\Lambda_{aa} + \Sigma_{ab}\Lambda_{ba} = \mathbb{I} \rightarrow \Lambda_{aa} = \Sigma_{aa}^{-1}(\mathbb{I} - \Sigma_{ab}\Lambda_{ba}) = \Sigma_{aa}^{-1} - \Sigma_{aa}^{-1}\Sigma_{ab}\Lambda_{ba}$

(2) $\Sigma_{aa}\Lambda_{ab} + \Sigma_{ab}\Lambda_{bb} = 0 \rightarrow \Lambda_{ab} = -\Sigma_{aa}^{-1}\Sigma_{ab}\Lambda_{bb}$

(3) $\Sigma_{ba}\Lambda_{aa} + \Sigma_{bb}\Lambda_{ba} = 0 \rightarrow \Lambda_{ba} = -\Sigma_{bb}^{-1}\Sigma_{ba}\Lambda_{aa}$

(4) $\Sigma_{ba}\Lambda_{ab} + \Sigma_{bb}\Lambda_{bb} = \mathbb{I} \rightarrow \Lambda_{bb} = \Sigma_{bb}^{-1} - \Sigma_{bb}^{-1}\Sigma_{ba}\Lambda_{ab}$

$$\Downarrow$$

$$\Lambda_{aa} = \Sigma_{aa}^{-1} - \Sigma_{aa}^{-1}\Sigma_{ab}\Lambda_{ba} = \Sigma_{aa}^{-1} - \Sigma_{aa}^{-1}\Sigma_{ab}(-\Sigma_{bb}^{-1}\Sigma_{ba}\Lambda_{aa})$$

$$\Lambda_{aa} - \Sigma_{aa}^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}\Lambda_{aa} = \Sigma_{aa}^{-1}$$

$$\Lambda_{aa}(\mathbb{I} - \Sigma_{aa}^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}) = \Sigma_{aa}^{-1}$$

$$\Lambda_{aa}(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}) = \mathbb{I}$$

$$\Lambda_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$$

Applying the same procedure:

$$\Lambda_{bb} = (\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}$$

$$\Lambda_{ba} = -\Sigma_{bb}^{-1}\Sigma_{ba}(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$$

$$\Lambda_{ab} = -\Sigma_{aa}^{-1}\Sigma_{ab}(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}$$

Referring to the textbook page #87 equation (2.76):

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix} = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$$

where : $M = (A - BD^{-1}C)^{-1}$

Continuing the derivations of Q (see page #5):

$$Q(x_a,x_b) = (x_a-\mu_a)^T\Lambda_{aa}(x_a-\mu_a) + (x_a-\mu_a)^T\Lambda_{ab}(x_b-\mu_b)$$
$$+ (x_b-\mu_b)^T\Lambda_{ba}(x_a-\mu_a) + (x_b-\mu_b)^T\Lambda_{bb}(x_b-\mu_b)$$

This is due to : $\Lambda_{ab} = \Lambda_{ba}^T$

$$= (x_a-\mu_a)^T\boxed{\Lambda_{aa}}(x_a-\mu_a) + 2(x_a-\mu_a)^T\boxed{\Lambda_{ab}}(x_b-\mu_b) + (x_b-\mu_b)^T\boxed{\Lambda_{bb}}(x_b-\mu_b)$$

$\downarrow$

Substituting $\Lambda_{aa}, \Lambda_{ab}, \Lambda_{ba}, \Lambda_{bb}$ with their new definitions:

$$Q(x_a,x_b) = (x_a-\mu_a)^T\left(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}\right)^{-1}(x_a-\mu_a)$$
$$-2(x_a-\mu_a)^T\left(\Sigma_{aa}^{-1}\Sigma_{ab}\left(\Sigma_{bb}-\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)^{-1}\right)(x_b-\mu_b)$$
$$+(x_b-\mu_b)^T\left(\Sigma_{bb}-\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)^{-1}(x_b-\mu_b)$$

→ this is different from

have the same term

$\downarrow$

The inverse matrix $(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$ or Λ_{aa} can be defined differently:

$$\Lambda_{aa} = \left(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}\right)^{-1} = \Sigma_{aa}^{-1} - \Sigma_{aa}^{-1}\Sigma_{ab}\Lambda_{ba}$$

→ see the previous page (#6)

$$= \Sigma_{aa}^{-1} - \Sigma_{aa}^{-1}\Sigma_{ab}\Lambda_{ab}^T = \Sigma_{aa}^{-1} + \Sigma_{aa}^{-1}\Sigma_{ab}\left(\Sigma_{aa}^{-1}\Sigma_{ab}\Lambda_{bb}\right)^T$$
$$= \Sigma_{aa}^{-1} + \Sigma_{aa}^{-1}\Sigma_{ab}\boxed{\Lambda_{bb}^T}\Sigma_{ab}^T\left(\Sigma_{aa}^{-1}\right)^T \quad ; \ \Lambda_{bb}^T = \Lambda_{bb} \ ; \ \Sigma_{ab}^T = \Sigma_{ba} \ ; \ \left(\Sigma_{aa}^{-1}\right)^T = \Sigma_{aa}^{-1}$$
$$= \Sigma_{aa}^{-1} + \Sigma_{aa}^{-1}\Sigma_{ab}\left(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)^{-1}\Sigma_{ba}\Sigma_{aa}^{-1}$$

$\Downarrow$

Hence:

$$Q(x_a,x_b) = (x_a-\mu_a)^T\left[\Sigma_{aa}^{-1} + \Sigma_{aa}^{-1}\Sigma_{ab}\left(\Sigma_{bb}-\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)^{-1}\Sigma_{ba}\Sigma_{aa}^{-1}\right](x_a-\mu_a)$$
$$-2(x_a-\mu_a)^T\left[\Sigma_{aa}^{-1}\Sigma_{ab}\left(\Sigma_{bb}-\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)^{-1}\right](x_b-\mu_b)$$
$$+(x_b-\mu_b)^T\left(\Sigma_{bb}-\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)^{-1}(x_b-\mu_b)$$

$$= (x_a-\mu_a)^T\Sigma_{aa}^{-1}(x_a-\mu_a)$$
$$+(x_a-\mu_a)^T\left[\Sigma_{aa}^{-1}\Sigma_{ab}\left(\Sigma_{bb}-\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)^{-1}\Sigma_{ba}\Sigma_{aa}^{-1}\right](x_a-\mu_a)$$
$$-2(x_a-\mu_a)^T\left[\Sigma_{aa}^{-1}\Sigma_{ab}\left(\Sigma_{bb}-\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)^{-1}\right](x_b-\mu_b)$$
$$+(x_b-\mu_b)^T\left(\Sigma_{bb}-\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)^{-1}(x_b-\mu_b)$$

$$Q(x_a, x_b) = (x_a-\mu_a)^T \Sigma_{aa}^{-1} (x_a-\mu_a)$$
$$+ (x_a-\mu_a)^T [\Sigma_{aa}^{-1} \Sigma_{ab} (\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1} \Sigma_{ba}\Sigma_{aa}^{-1}] (x_a-\mu_a)$$
$$- 2(x_a-\mu_a)^T [\Sigma_{aa}^{-1} \Sigma_{ab} (\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}] (x_b-\mu_b)$$
$$+ (x_b-\mu_b)^T (\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1} (x_b-\mu_b)$$

Let's define :

$$w_a^T = (x_a-\mu_a)^T \Sigma_{aa}^{-1}\Sigma_{ab} \rightarrow w_a = \Sigma_{ab}\Sigma_{aa}^{-1}(x_a-\mu_a)$$
$$w_b = (x_b-\mu_b)$$
$$M = (\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}$$

$$\Downarrow$$

$$Q(x_a, x_b) - (x_a-\mu_a)^T \Sigma_{aa}^{-1}(x_a-\mu_a)$$
$$= w_a^T M w_a - 2 w_a^T M w_b + w_b^T M w_b$$
$$= w_a^T M w_a - w_a^T M w_b - w_a^T M w_b + w_b^T M w_b$$
$$= w_a^T M (w_a - w_b) - (w_a-w_b)^T M w_b$$
$$= w_a^T M (w_a - w_b) - w_b^T M (w_a - w_b)$$
$$= (w_a-w_b)^T M (w_a-w_b) = (w_b-w_a)^T M (w_b-w_a)$$

$$Q(x_a, x_b) = (x_a-\mu_a)^T \Sigma_{aa}^{-1}(x_a-\mu_a) + (w_b-w_a)^T M (w_b-w_a)$$

$$\downarrow$$

$$Q(x_a, x_b) = Q_1(x_a) + Q_2(x_a, x_b)$$

$$Q_1(x_a) = (x_a-\mu_a)^T \Sigma_{aa}^{-1}(x_a-\mu_a) \qquad Q_2(x_a,x_b) = (w_b-w_a)^T M (w_b-w_a)$$

Recall : $p(x) = p(x_a, x_b) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(Q(x_a,x_b)\right)$

What is $|\Sigma|$?

$\det(\Sigma)$

$$|\Sigma| = \begin{vmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{vmatrix} = \Sigma_{aa}\Sigma_{bb} - \Sigma_{ab}\Sigma_{ba} \quad ; \quad \Sigma = \begin{bmatrix} \Sigma_{aa} & 0 \\ \Sigma_{ba} & I \end{bmatrix} \begin{bmatrix} I & \Sigma_{aa}^{-1}\Sigma_{ab} \\ 0 & \Sigma_{bb} - \Sigma_{ab}^T\Sigma_{aa}^{-1}\Sigma_{ab} \end{bmatrix}$$

$$|\Sigma| = |\Sigma_a||\Sigma_b| \qquad = \Sigma_A \Sigma_B$$

$$\frac{1}{|\Sigma|} = \frac{1}{|\Sigma_A|}\frac{1}{|\Sigma_B|} = \frac{1}{|\Sigma_{aa}|} \cdot \frac{1}{|\Sigma_{bb} - \Sigma_{ab}^T \Sigma_{aa}^{-1}\Sigma_{ab}|} = \frac{1}{|\Sigma_{aa}|} \cdot \frac{1}{|M^{-1}|}$$

$$Q(x_a, x_b) = (x_a-\mu_a)^T \Sigma_{aa}^{-1} (x_a-\mu_a) + (w_b - w_a)^T M (w_b - w_a)$$

$$p(x_a, x_b) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_{aa}|} \frac{1}{|M^{-1}|} \exp\left(-\frac{1}{2} Q(x_a, x_b)\right) = p(x_b|x_a)\, p(x_a)$$

$$= \underbrace{\frac{1}{(2\pi)^{P/2}} \frac{1}{|\Sigma_{aa}|} \exp\left(-\frac{1}{2}(x_a-\mu_a)^T \Sigma_{aa}^{-1} (x_a-\mu_a)\right)} \underbrace{\frac{1}{(2\pi)^{Q/2}} \frac{1}{|M^{-1}|} \exp\left(-\frac{1}{2} (w_b-w_a)^T M (w_b-w_a)\right)}$$

$$p(x_a) = G(x_a; \mu_a, \Sigma_{aa})$$

What are w_a and w_b actually.

Recall : $w_a = \Sigma_{ab}^T \Sigma_{aa}^{-1} (x_a - \mu_a)$

$w_b = x_b - \mu_b$

$$w_b - w_a = (x_b - \mu_b) - \Sigma_{ba} \Sigma_{aa}^{-1} (x_a - \mu_a)$$

Hence: $p(x_b | x_a) = \frac{1}{(2\pi)^{Q/2}} \frac{1}{|M^{-1}|} \exp\left(-\frac{1}{2}(x_b - b) M (x_b - b)\right)$

where : $b = \mu_b + \Sigma_{ba} \Sigma_{aa}^{-1} (x_a - \mu_a)$

$$p(x_b | x_a) = G(x_b ; b, M^{-1})$$

Therefore :

$$\boxed{\begin{aligned} p(x_a, x_b) &= p(x_a)\, p(x_b | x_a) \\ &= G(x_a ; \mu_a, \Sigma_{aa})\, G(x_b ; b, M_b^{-1}) \\ \text{where: } & \\ b &= \mu_b + \Sigma_{ba} \Sigma_{aa}^{-1} (x_a - \mu_a) \\ M_b &= (\Sigma_{bb} - \Sigma_{ba} \Sigma_{aa}^{-1} \Sigma_{ab})^{-1} \end{aligned}}$$

We successfully solve the integra-tion problem analytically !!

We know also that : $p(x_a, x_b) = p(x_b)\, p(x_a | x_b)$

$$= G(x_b ; \mu_b, \Sigma_{bb})\, G(x_a ; a, M_a^{-1})$$

where : $a = \mu_a + \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)$

$M_a = (\Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba})^{-1}$

[5] Conditional & Marginal Gaussian Distributions: CASE 2

In the previous case (case 1), we assume: $p(\bar{x}) = p(x_a, x_b) = G(\bar{x}; \bar{\mu}, \Sigma)$

Thus, $\bar{x}_a$ & $\bar{x}_b$ are part of a single Gaussian.

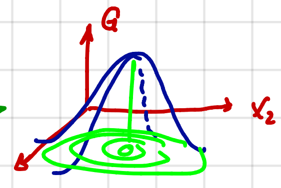
$\Downarrow$

Now, imagine we have two variables $\bar{x}$ and $\bar{y}$, and both are not part of the same Gaussian:

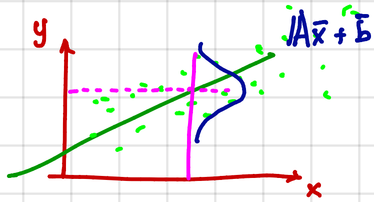
$$\underset{D\times 1}{\bar{y}} = \underset{D\times M}{A}\,\underset{M\times 1}{\bar{x}} + \underset{D\times 1}{\bar{b}}$$

where:

Marginal distribution $\rightarrow$ $p(\bar{x}) = G(\bar{x}; \bar{\mu}, \Lambda^{-1})$



Conditional distribution $\rightarrow$ $p(\bar{y}|\bar{x}) = G(\bar{y}; A\bar{x} + \bar{b}, L^{-1})$



$\Downarrow$

Questions: $p(\bar{x}|\bar{y}) = ?$

Full Bayesian Inference: $\boxed{p(\bar{x}|\bar{y})} = \dfrac{p(y|x)\,p(x)}{\boxed{p(y)}}$ $\rightarrow$ $p(y) = ?$

$\Downarrow$

In our regression problem, the above statements can be interpreted as:

$$y_n = \sum_{m=1}^{M} w_m \phi_m(x_n) + w_0 \quad ; \quad t_n = y_n + \varepsilon$$

Assumptions:

$p(\bar{w}) = G(\bar{w}; \bar{\mu}, \Lambda^{-1})$ $\rightarrow$ Prior

$p(t_n|\bar{w}) = G(t_n; y_n, L^{-1})$ $\rightarrow$ Likelihood

$\downarrow$

Questions: $p(\bar{w}|t_n) = ?$ $\rightarrow$ Conditional Gaussian distribution

$p(\bar{w}|t_n) = \dfrac{p(t_n|\bar{w})\,p(\bar{w})}{p(t_n)}$ $\rightarrow$ $p(t_n) = ?$ $\rightarrow$ Marginal Gaussian distribution.

Solution:

$$\bar{z} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} \rightarrow \log p(\bar{z}) = \log p(\bar{x}, \bar{y}) = \log p(\bar{y} | \bar{x}) + \log p(\bar{x})$$

$$= -\frac{1}{2} (\bar{x} - \bar{\mu})^T \Lambda (\bar{x} - \bar{\mu}) - \frac{1}{2} (y - A\bar{x} - \bar{b})^T L (y - A\bar{x} + \bar{b}) + \text{Const}$$

$\downarrow$

This equation, w.r.t y, is a quadratic function.

Hence, it's likely $p(y) = \int p(y, x)\, dx$ is a Gaussian.

[there is something special with the Gaussian integral. It might produce another Gaussian]

$\Downarrow$

Our task is then to compute μ_y & Σ_y : $G(\bar{y}; \mu_y, \Sigma_y)$

$$\log p(\bar{z}) = -\frac{1}{2} (\bar{x} - \bar{\mu})^T \Lambda (\bar{x} - \bar{\mu}) - \frac{1}{2} (y - A\bar{x} - \bar{b})^T L (y - A\bar{x} + \bar{b}) + \text{Const}$$

$$-2 \log p(z) - \text{Const} = (\bar{x} - \bar{\mu})^T \Lambda (\bar{x} - \bar{\mu}) + (y - A\bar{x} - \bar{b})^T L (y - A\bar{x} + \bar{b})$$

$$\begin{aligned} = \; & \bar{x}^T \Lambda \bar{x} + 2\bar{x}^T \Lambda \bar{\mu} + \bar{\mu}^T \Lambda \bar{\mu} \\ & + \bar{y}^T L \bar{y} - \bar{y}^T L A\bar{x} - y^T L \bar{b} \\ & - (A\bar{x})^T L \bar{y} + (A\bar{x})^T L A\bar{x} + (A\bar{x})^T L \bar{b} \\ & - \bar{b}^T L \bar{y} + \bar{b}^T L A\bar{x} + \bar{b}^T L \bar{b} \end{aligned}$$

$$\begin{aligned} = \; & \bar{x}^T \Lambda \bar{x} + \bar{y}^T L \bar{y} - \bar{y}^T L A\bar{x} - \bar{x}^T A^T L \bar{y} + \bar{x}^T A^T L A\bar{x} \rightarrow \text{Quadratic w.r.t. } \bar{x} \,\&\, \bar{y} \\ & - 2\bar{x}^T \Lambda \mu - 2\bar{y}^T L \bar{b} + 2\bar{x}^T A^T L \bar{b} \rightarrow \text{linear w.r.t. } \bar{x} \,\&\, \bar{y} \\ & + \bar{\mu}^T \Lambda \mu + \bar{b}^T L \bar{b} \rightarrow \text{constant} \end{aligned}$$

[•] Focusing on the quadratic terms (to compute Σ_y):

$$\bar{x}^T \Lambda \bar{x} + \bar{y}^T L \bar{y} - \bar{y}^T L A \bar{x} - \bar{x}^T A^T L \bar{y} + \bar{x}^T A^T L A \bar{x}$$

$$-\frac{1}{2}\left(\underline{\bar{x}^T \Lambda \bar{x}} + \bar{y}^T L \bar{y} - \bar{y}^T L A \bar{x} - \bar{x}^T A^T L \bar{y} + \underline{\bar{x}^T A^T L A \bar{x}}\right)$$

$$= -\frac{1}{2}\left[\bar{x}^T (\Lambda + A^T L A)\bar{x} - \bar{x}^T A^T L \bar{y} - \bar{y}^T L A \bar{x} + \bar{y}^T L \bar{y}\right]$$

$$= -\frac{1}{2}\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix}^T \underbrace{\begin{bmatrix}\Lambda + A^T L A & -A^T L\\ -LA & L\end{bmatrix}}_{} \begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix}$$

Let's define : inverse covariance matrix = R

Inverse of partition matrix (see page #4)

$$\mathrm{cov}(z) = R^{-1} = \begin{bmatrix}\Lambda^{-1} & \Lambda^{-1}A^T\\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T\end{bmatrix}$$

$$\mathrm{cov}\begin{bmatrix}\bar{x}\\ \bar{y}\end{bmatrix} \rightarrow \begin{array}{l}\mathrm{cov}(\bar{x}) = \Lambda^{-1}\\ \mathrm{cov}(\bar{y}) = L^{-1} + A\Lambda^{-1}A^T\end{array}$$

$\Rightarrow$ The covariance : Σ_y

[•] Focusing on the linear terms (to compute μ_y)

$$\bar{x}^T \Lambda \bar{\mu} - \bar{x}^T A^T L \bar{b} + \bar{y}^T L \bar{b} = \begin{pmatrix}x\\ y\end{pmatrix}^T \begin{pmatrix}\Lambda\bar{\mu} - A^T L \bar{b}\\ L\bar{b}\end{pmatrix}$$

The Gaussian format is: $(z-\mu_z)^T \Lambda_z (z-\mu_z)$

$$z^T \Lambda_z z - 2 z^T \Lambda \mu_z + \mu_z^T \Lambda \mu_z$$

the linear term $\begin{pmatrix}x\\ y\end{pmatrix}^T \Lambda_z \mu_z$

$$\Lambda_z \mu_z = \begin{pmatrix}\Lambda\bar{\mu} - A^T L \bar{b}\\ L\bar{b}\end{pmatrix}$$

$$\Lambda_z = \mathrm{cov}(z) = R$$

$$\mu_z = R^{-1}\begin{pmatrix}\Lambda\bar{\mu} - A^T L \bar{b}\\ L\bar{b}\end{pmatrix}; \quad \text{Recall: } R^{-1} = \begin{pmatrix}\Lambda^{-1} & \Lambda^{-1}A^T\\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T\end{pmatrix}$$

$$\mu_z = \begin{pmatrix}\mu_x\\ \mu_y\end{pmatrix} = \begin{pmatrix}\Lambda^{-1} & \Lambda^{-1}A^T\\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T\end{pmatrix}\begin{pmatrix}\Lambda\bar{\mu} - A^T L \bar{b}\\ L\bar{b}\end{pmatrix}$$

$$= \begin{pmatrix}\mu - \Lambda^{-1}\cancel{A^T L}\bar{b} + \Lambda^{-1}\cancel{A^T L}\bar{b}\\ A\bar{\mu} - A\cancel{\Lambda^{-1}A^T}Lb + \bar{b} + A\cancel{\Lambda^{-1}A^T}Lb\end{pmatrix} = \begin{pmatrix}\mu\\ A\bar{\mu} + b\end{pmatrix}$$

$$\mu_y = A\bar{\mu} + b$$

Hence : $\boxed{p(y) = G(\bar{y}; A\bar{\mu} + b, L^{-1} + A\Lambda^{-1}A^T)}$ $\rightarrow$ We solved the integration problem analytically!

[•] How about $p(x|y) = ?$ (the conditional Gaussian distribution)

Recall:

$$p(x_a|x_b) = \frac{p(x_b|x_a)\,p(x_a)}{p(x_b)} = \frac{G(x_b; \mu_{b|a}, \Sigma_{b|a})\, G(x_a; \mu_a, \Sigma_{aa})}{G(x_b; \mu_b, \Sigma_{bb})}$$

$$= G(x_a; \mu_{a|b}, \Sigma_{a|b})$$

where: $\mu_{a|b} = \Lambda_{aa}^{-1}(\Lambda_{aa}\mu_a - \Lambda_{ab}(x_b - \mu_b)) = \mu_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b - \mu_b)$

$\Sigma_{a|b} = \Lambda_{aa}^{-1} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$

Our current problem: $p(\bar{x}|\bar{y})$, which is similar to the above form:

$$p(\bar{x}|\bar{y}) = \frac{p(\bar{y}|\bar{x})\,p(\bar{x})}{p(\bar{y})} = \frac{G(\bar{y}; \bar{\mu}_{y|x}, L^{-1})\, G(\bar{x}; \bar{\mu}_x, \Lambda^{-1})}{G(\bar{y}; \bar{\mu}_y, \Sigma_y)}$$

Hence: $p(\bar{x}|\bar{y}) = G(\bar{x}; \bar{\mu}_{x|y}, \Sigma_{x|y})$

where:

$\bar{\mu}_{x|y} = \Lambda_{aa}^{-1}(\Lambda_{aa}\mu_x - \Lambda_{ab}(y - \mu_y))$

$\Sigma_{x|y} = \Lambda_{aa}^{-1}$

Note: R in our previous notation is Λ.

!

and $\bar{\mu}_x = \bar{\mu}$; $\bar{\mu}_y = \mathbb{A}\mu + \bar{b}$

Recall (page #11):

$$R = \begin{bmatrix} \Lambda + \mathbb{A}^T L \mathbb{A} & -\mathbb{A}^T L \\ -L\mathbb{A} & L \end{bmatrix}$$

$\Downarrow$

$\Sigma_{x|y} = (\Lambda + \mathbb{A}^T L \mathbb{A})^{-1}$

$\bar{\mu}_{x|y} = (\Lambda + \mathbb{A}^T L \mathbb{A})^{-1}([\Lambda + \mathbb{A}^T L \mathbb{A}]\mu + \mathbb{A}^T L[\bar{y} - \mathbb{A}\mu - \bar{b}])$

$= (\Lambda + \mathbb{A}^T L \mathbb{A})^{-1}(\Lambda\mu + \cancel{\mathbb{A}^T L \mathbb{A}\mu} + \mathbb{A}^T L(\bar{y} - \bar{b}) - \cancel{\mathbb{A}^T L \mathbb{A}\mu})$

$$\boxed{\begin{aligned} \mu_{x|y} &= (\Lambda + \mathbb{A}^T L \mathbb{A})^{-1}(\mathbb{A}^T L(\bar{y} - \bar{b}) + \Lambda\bar{\mu}) \\ \Sigma_{x|y} &= (\Lambda + \mathbb{A}^T L \mathbb{A})^{-1} \end{aligned}}$$

See textbook page #93.

[6] Summary

CASE 1:

$$\bar{x} = \begin{pmatrix} \bar{x}_a \\ \bar{x}_b \end{pmatrix} \text{ and } p(\bar{x}) = G(\bar{x}; \mu, \Sigma)$$
$$= G\left(\begin{bmatrix} \bar{x}_a \\ \bar{x}_b \end{bmatrix} ; \begin{bmatrix} \bar{\mu}_a \\ \bar{\mu}_b \end{bmatrix} , \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \right)$$

Marginal Gaussian distribution:

$$p(\bar{x}_a) = G(\bar{x}_a ; \mu_a , \Sigma_{aa})$$
$$p(\bar{x}_b) = G(\bar{x}_b ; \mu_b , \Sigma_{bb})$$

Conditional Gaussian distribution:

$$p(\bar{x}_a | \bar{x}_b) = G(\bar{x}_a ; \mu_{a|b} , \Sigma_{a|b}^{-1})$$

where:

$$\mu_{a|b} = \mu_a + \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)$$
$$\Sigma_{a|b} = (\Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba})^{-1}$$

$$p(\bar{x}_b | x_a) = G(\bar{x}_b ; \mu_{b|a} , \Sigma_{b|a}^{-1})$$

where:

$$\mu_{b|a} = \mu_b + \Sigma_{ba} \Sigma_{aa}^{-1} (x_a - \mu_a)$$
$$\Sigma_{b|a} = (\Sigma_{bb} - \Sigma_{ba} \Sigma_{aa}^{-1} \Sigma_{ab})^{-1}$$

CASE 2:

$$\bar{y} = \bar{A}\bar{x} + \bar{b}$$

where:

$$p(\bar{x}) = G(\bar{x} ; \bar{\mu}, \Lambda^{-1}) \rightarrow \text{Prior}$$
$$p(\bar{y}|\bar{x}) = G(\bar{y} ; A\bar{x} + b, L^{-1}) \rightarrow \text{Likelihood}$$

Marginal Gaussian distribution:

$$p(\bar{y}) = G(\bar{y} ; A\bar{\mu} + \bar{b} , L^{-1} + A\Lambda^{-1}A^T)$$

marginalization

Conditional Gaussian distribution:

$$p(\bar{x}|\bar{y}) = G(\bar{x} ; \bar{\mu}_{x|y} , \Sigma_{x|y})$$

posterior

where:

$$\mu_{x|y} = (\Lambda + A^T L A)^{-1} (A^T L (\bar{y} - \bar{b}) + \Lambda\bar{\mu})$$
$$\Sigma_{x|y} = (\Lambda + A^T L A)^{-1}$$